

WEB APPENDIX

Abbreviations: HCHS/SOL, Hispanic Community Health Study/Study of Latinos; DLW, doubly labeled water; TEE, total energy expenditure; 24 HR, 24 hour dietary recall.

Collaborating centers

Study procedures were approved by the institutional review boards of all the study sites and coordinating/reading centers.

Recruitment

Study of Latinos: Nutrition & Physical Activity Study staff sent letters describing the study to potential participants who were within approximately six months from their baseline visit in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) parent study for all sites, except San Diego which was allowed a window of twelve months because of logistical issues in recruitment. The letters were followed by a telephone screening call from study team members. All eligible and interested participants were asked to provide their written informed consent. Participants received \$25.00 for attending the first visit, an additional \$75.00 upon study completion, and were reimbursed for transportation costs. Lunch was provided at the first visit and a snack at the second visit.

Doubly labeled water (DLW) procedure

The DLW mixture was pre-mixed by the stable isotope manufacturer and dispensed in 1-L leak-proof plastic bottles. At the time of a study, the staff at the study sites calculated the amount of DLW that is needed for each study participant as follows:

Amount of DLW needed (g) = 1.4 g of DLW mixture / kg body wt x body wt (kg)

For five participants, no plasma was collected and therefore Total Energy Expenditure (TEE) from the spot urines was used.

Respiratory Quotient for calculation of TEE from DLW

We created an ethnic-specific food quotient (1) based on the 24 dietary recall (24 HR) in the HCHS/SOL parent study and calculated energy values using both the standard respiratory quotient for Western populations (0.86) and an ethnic specific one. However, the difference in energy values using either method was minimal. The difference in these calculated food quotient values varied from 0.8566 to 0.86921, or a difference of 0.01 unit. An increase of 0.01 unit of the food quotient would increase TEE by only 1%. Therefore, we did not expect much difference in TEE whether we assumed a respiratory quotient value of 0.86 or used the ethnic-specific food quotient. Also, given that the 24HR are based on self-report, we opted for using the standard respiratory quotient value of 0.86 for estimating TEE from DLW.

Assessment of urinary completion

We assessed para-amino benzoic acid recovery in the 24 hr urine sample to check completion of urinary collection on a random sample consisting of 10% of the participants. For <70% para-amino benzoic acid recovery, samples from female participants indicated 68.45 g protein /day (95% CI, 58.01, 80.77); males 101.93 g protein/day, (95% CI, 72.19, 143.91). For para-amino benzoic acid recovery that was $\geq 70\%$, samples from females indicated 79.34 g protein/day, (95% CI, 64.41, 97.74); males 102.74 g protein/day (95% CI, 82.05, 128.66). We also performed sensitivity analyses on observations with extremely low urinary volume (below 10th percentile), reflecting potential incompleteness of the 24-hr urine collection, to check their influence on the regression calibration coefficients. Results were similar, while the strength of evidence for language preference was weakened slightly.

Statistical analysis

Calibration equations for use in disease risk association studies were developed using generalized estimating equations that predicted true intakes of energy and protein given the self-reported intakes and data on study subject characteristics based on the following measurement error models (2,3). For the regression calibration we assumed that the biomarker assessment W (DLW for energy and urinary nitrogen for protein) in log format adheres to a classical measurement model,

$$W = Z + e \quad (A)$$

where Z represents the nutrient of interest such as energy consumption, protein intake or protein density or its log-transformed values, and e is an error term that is assumed to be independent of Z and other study subject characteristics. Z can be regarded as the logarithm of average daily consumption for the nutrient of interest over a fairly short period of time, such as 6-12 months in proximity to the biomarker data collection period. Q represents an estimate of the self-reported nutrient in this case energy, protein and protein density data from the 24 hr dietary recalls. In order to incorporate participant characteristics that can influence reporting, the following expanded measurement model was considered for the (log-transformed) Q :

$$Q = S_0 + S_1Z + S_2V + S_3VZ + r + u \quad (B)$$

where, S_0 , S_1 , S_2 , and S_3 are regression parameters to be estimated, V is a set of participant characteristics that may relate to systematic bias in the assessment (such as age, body mass index (BMI), and background), r is a person-specific error variable, and u is an independent measurement error term. Also, r and u are independent of Z , V , and e .

Dietary variables were entered into a linear regression model in addition to other participant characteristics, including age, body mass index, ethnicity and other factors such as smoking, to identify associations with the difference between log (self-reported nutrient) and log (biomarker). Finally, we conducted a series of linear regressions of log (biomarker) on log (self-report nutrient) and participant characteristics, using backwards stepwise selection keeping only those characteristics with $p < 0.10$. Based on our measurement error models these calibration equations allow estimation of targeted nutritional value Z based on Q and V . We calculated the fraction of the total variance in the log-transformed biomarker (R^2) that could be explained by the self-report assessment and participant characteristics. We also calculated the adjusted R^2 values as R^2 value divided by the correlation between the repeat measure errors for the biomarker W (3). The adjusted R^2 can be interpreted as the percentage of variation in Z explained by Q and V in the calibration model.

Web Table 1. Sensitivity analysis for regression calibration coefficients for log-transformed biomarker, where the log-transformed self-report values are based on the 24HR recall 2-day mean and individuals with urine volumes below the 10th percentile were excluded ($n = 423$), Study of Latinos: Nutrition and Physical Activity Assessment Study, 2010-2012

	Protein			Protein Density		
	β	SE	P Value ^a	β	SE	P Value ^a
Intercept	4.7562	0.035	<0.001	2.8167	0.026	<0.001
24HR Recall 2-Day Mean ^b	0.1208	0.037	0.001	0.2581	0.063	<0.001
Age ^b , yr	.	.	.	0.0017	0.001	0.18
BMI ^b , kg/m ²	0.0155	0.002	<0.001	.	.	.
Female	-0.2216	0.032	<0.001	.	.	.
Dominican	-0.0414	0.052	0.425	-0.0266	0.053	0.617
Central American	-0.0955	0.044	0.029	-0.0577	0.043	0.184
Cuban	-0.0566	0.043	0.19	0.0132	0.045	0.771
Puerto Rican	-0.138	0.04	0.001	-0.129	0.042	0.002
South American	0.0604	0.047	0.197	0.1186	0.048	0.014
English Preference	-0.0838	0.039	0.032	-0.0608	0.046	0.184
Income Missing
Income \$10,001-\$20,000
Income \$20,001-\$40,000
Income \$40,001-\$50,000
Income \$50,001-\$75,000
Income >\$75,000
Current Smoker	-0.1213	0.041	0.003	-0.167	0.042	<0.001

Abbreviations: SE, standard error; BMI, body mass index.

^a Overall P -values for ethnicity were significant for protein ($P<0.001$) and protein density ($P<0.001$).

^b Age centered on mean, 46 years; BMI centered on mean, 29.6 kg/m²; log-transformed 2-day mean of protein centered on 4.285903; log-transformed 2-day mean of protein density centered on 2.80012.

Web Table 2. *R*-squared and adjusted *R*-squared values for the covariates in each of the regression models in Table 5, Study of Latinos: Nutrition and Physical Activity Assessment Study, 2010-2012^a

	ENERGY Total $R^2 = 0.54$ Adj $R^2 = 0.67$		PROTEIN Total $R^2 = 0.26$ Adj $R^2 = 0.39$		PROTEIN DENSITY Total $R^2 = 0.16$ Adj $R^2 = 0.28$	
	R^2	Adjusted R^2	R^2	Adjusted R^2	R^2	Adjusted R^2
Intercept						
24HR 2-day mean	0.079	0.097	0.044	0.066	0.032	0.054
Age, yr	0.023	0.028			0.01	0.017
BMI, kg/m ²	0.125	0.155	0.055	0.083		
Female	0.268	0.331	0.08	0.122		
Ethnicity	0.027	0.033	0.03	0.045	0.044	0.075
English preference			0.017	0.026	0.042	0.072
Income	0.021	0.026				
Current smoker	0.002	0.002	0.033	0.050	0.037	0.063

^a R^2 values for individual covariates were calculated by the regressions of that covariate alone and then rescaled to add up to total R^2 for the main model. Note: for categorical variables, the R^2 value for the model includes each of the subcategory variables. Regressions with SOLNAS main study data (not including reliability data) were used to produce traditional R^2 coefficients.

Web Table 3. Regression calibration coefficients for log-transformed biomarker for females, Study of Latinos: Nutrition and Physical Activity Assessment Study, 2010-2012

	Energy			Protein			Protein Density		
	β	SE	<i>P</i> Value ^a	β	SE	<i>P</i> Value ^a	β	SE	<i>P</i> Value ^a
Intercept	7.7128	0.024	<0.001	4.5152	0.028	<0.001	2.7958	0.03	<0.001
24HR Recall 2Day Mean ^b	0.036	0.022	0.101	0.078	0.048	0.104	0.2314	0.073	0.001
Age ^b , yr	-0.001	0.001	0.085	.	.	.	0.0032	0.002	0.038
BMI ^b , kg/m ²	0.0151	0.001	<0.001	0.0145	0.003	<0.001	.	.	.
Dominican	-0.0629	0.031	0.04	-0.1184	0.062	0.057	-0.0708	0.062	0.255
Central American	-0.0756	0.028	0.007	-0.1077	0.061	0.075	-0.0407	0.058	0.483
Cuban	-0.0099	0.027	0.714	-0.0516	0.054	0.337	-0.038	0.057	0.508
Puerto Rican	-0.0169	0.02	0.398	-0.1436	0.047	0.002	-0.1542	0.055	0.005
South American	-0.0786	0.029	0.006	0.0162	0.056	0.773	0.0997	0.061	0.103
English Preference	.	.	.	-0.0807	0.047	0.088	-0.0416	0.063	0.509
Income Missing	-0.0251	0.034	0.459
Income \$10,001- \$20,000	-0.0041	0.025	0.869
Income \$20,001- \$40,000	-0.0104	0.027	0.703
Income \$40,001- \$50,000	-0.0049	0.036	0.893
Income \$50,001- \$75,000	-0.0115	0.041	0.779
Income >\$75,000	-0.0635	0.084	0.447
Current Smoker	0.0425	0.019	0.028	-0.1393	0.047	0.003	-0.1845	0.052	<0.001

Abbreviations: SE, standard error; BMI, body mass index.

^a Overall *P*-values for ethnicity were significant for energy (*P*=0.011), protein (*P*=0.021), and protein density (*P*=0.010).

^b Age centered on mean, 46 years; BMI centered on mean, 29.6 kg/m²; log-transformed 2-day mean of energy centered on 7.489026; log-transformed 2-day mean of protein centered on 4.285903; log-transformed 2-day mean of protein density centered on 2.80012.

Web Table 4. Regression calibration coefficients for log-transformed biomarker for males, Study of Latinos: Nutrition and Physical Activity Assessment Study, 2010-2012

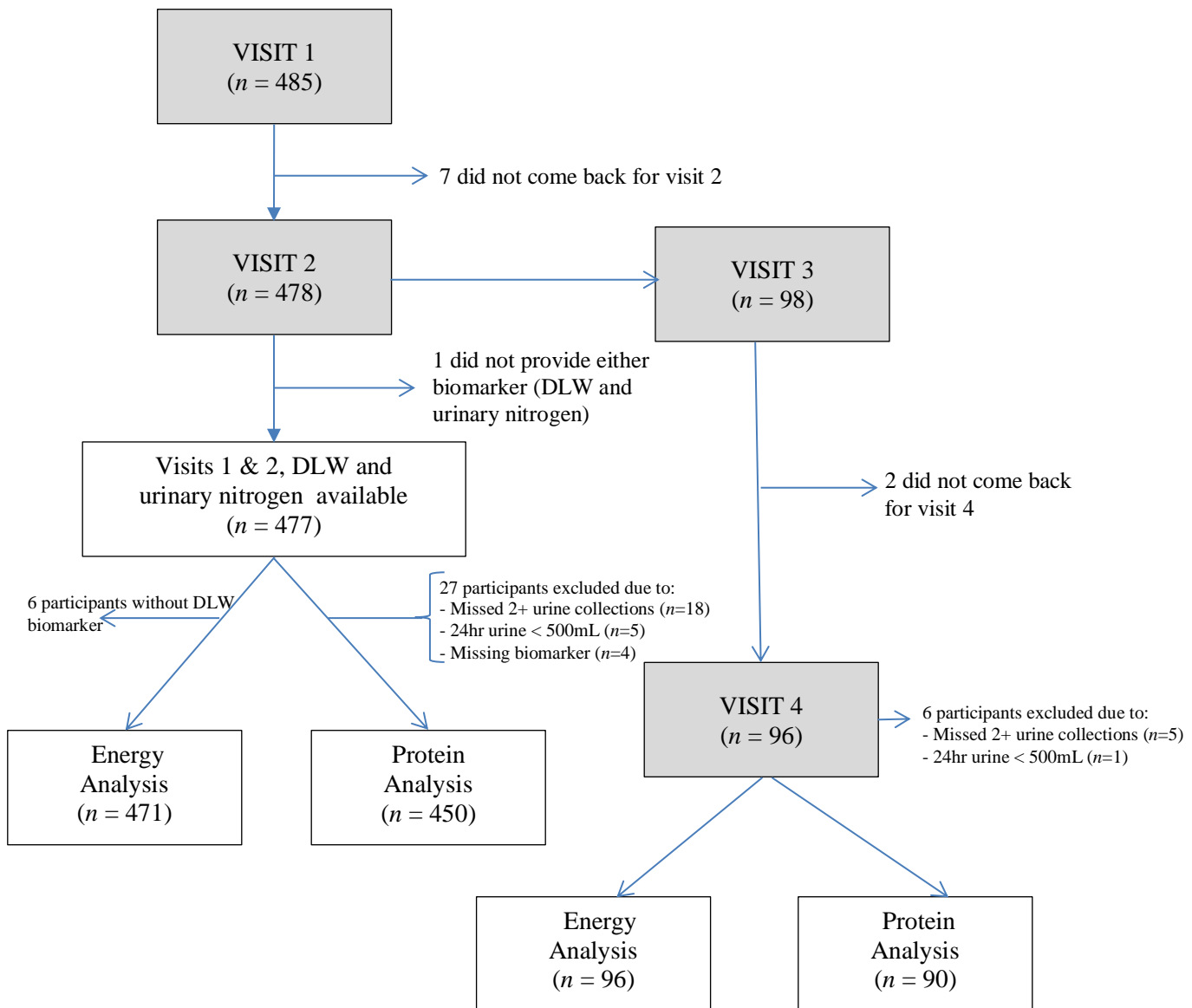
	Energy			Protein			Protein Density		
	β	SE	<i>P</i> Value ^a	β	SE	<i>P</i> Value ^a	β	SE	<i>P</i> Value ^a
Intercept	7.8903	0.049	<0.001	4.6908	0.055	<0.001	2.7607	0.059	<0.001
24HR Recall 2Day Mean ^b	0.067	0.027	0.014	0.1996	0.058	0.001	0.3154	0.104	0.002
Age ^b , yr	-0.0021	0.001	0.012	.	.	.	0.0033	0.002	0.123
BMI ^b , kg/m ²	0.0153	0.002	<0.001	0.0204	0.004	<0.001	.	.	.
Dominican	-0.019	0.043	0.661	0.1093	0.1	0.276	0.0789	0.111	0.475
Central American	-0.0281	0.044	0.52	-0.0786	0.072	0.273	-0.0621	0.081	0.444
Cuban	-0.1042	0.031	0.001	-0.062	0.08	0.436	0.0469	0.087	0.589
Puerto Rican	-0.0194	0.031	0.528	-0.0842	0.078	0.28	-0.0595	0.079	0.454
South American	-0.0498	0.034	0.14	0.1453	0.088	0.097	0.1919	0.088	0.03
English Preference	.	.	.	-0.173	0.066	0.009	-0.1378	0.076	0.069
Income Missing	0.0422	0.048	0.377
Income \$10,001- \$20,000	0.0187	0.046	0.682
Income \$20,001- \$40,000	0.0908	0.046	0.047
Income \$40,001- \$50,000	0.0611	0.054	0.262
Income \$50,001- \$75,000	0.0639	0.058	0.27
Income >\$75,000	-0.0818	0.059	0.166
Current Smoker	0.0509	0.028	0.07	-0.0947	0.071	0.183	-0.1387	0.071	0.05

Abbreviations: SE, standard error; BMI, body mass index.

^a Overall *P*-values for ethnicity were significant for energy (*P*=0.026), borderline significant for protein (*P*=0.055), and significant for protein density (*P*=0.046).

^b Age centered on mean, 46 years; BMI centered on mean, 29.6 kg/m²; log-transformed 2-day mean of energy centered on 7.489026; log-transformed 2-day mean of protein centered on 4.285903; log-transformed 2-day mean of protein density centered on 2.80012.

Web Figure 1. Flow chart of enrolled study subjects from enrollment to completion of study, Study of Latinos: Nutrition and Physical Activity Assessment Study, 2010-2012



REFERENCES

1. Black AE, Prentice AM, Coward WA. Use of food quotients to predict respiratory quotients for the doubly-labelled water method of measuring energy expenditure. *Hum Nutr Clin Nutr.* 1986;40(5):381-391.
2. Mossavar-Rahmani Y, Tinker LF, Huang Y, et al. Factors relating to eating style, social desirability, body image and eating meals at home increase the precision of calibration equations correcting self-report measures of diet using recovery biomarkers: findings from the Women's Health Initiative. *Nutr J.* 2013;12(1):63.
3. Neuhouser ML, Tinker L, Shaw PA, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol.* 2008;167(10):1247-1259.